Introduction to the Machine Learning and Predictive Analytics in Accounting, Finance and Management Minitrack

Peter Sarlin
Silo.AI
RiskLab at Arcada and Hanken School of Economics
peter@silo.ai

J´ozsef Mezei
Åbo Akademi University
Avaintec Oy, Helsinki
jmezei@abo.fi

Predictive analytics, making use of statistical modelling, machine learning and network analysis, has become one of the cornerstones of a modern organization in recent years. In particular, the applications to problems in different areas of business and economics play an in-creasing role in academic literature. The development and accessibility of innovative machine learning techniques in combination with the availability of (public) data-sets foster the growth of this domain. In the minitrack, the goal is to provide an avenue for sharing novel results on the development and application of novel machine learning techniques as used for (large) data-sets relevant to accounting, finance and management. Some of the re-cent developments present in this years' articles include deep learning, network analysis for cryptocurrencies, up-lift modeling for marketing and natural language processing.

This year we have received similar number of submis-sions to the minitrack as last year. This indicates the continued general interest in and recognition of the importance of the covered topics. After the review process, six articles have been selected to be presented in the conference. The accepted papers are representative of the various application domains (accounting, finance and marketing) and techniques (network analysis, text sentiment analysis and advanced machine learning models including neural networks and random forests) that are in the core of business and predictive analytics.

The first article "Accounting method selection using neural networks and multi-criteria decision making" is authored by Yang Duan, Chung-Hsing Yeh and David L. Dowe (Monash University). The authors combine neural networks and fuzzy linguistic multi-criteria decision models in a novel approach to aid decision makers in optimizing their accounting processes. In the work, the Full Cost and the Successful Effort accounting approaches are analyzed with the aim of selecting the optimal one. As the authors illustrate in an empirical analysis through the case of a multinational integrated oil and gas company, accounting methods can be analyzed and optimized in a rational and informed manner.

The second article "Follow the money: Revealing risky nodes in a Ransomware-Bitcoin network" by Adam Brian Turner, Stephen McCombie and Allon J. Uhlmann (Macquarie University) investigates the use of network analysis to model attacks in cryptocurrency networks. The authors focus on understanding ransomware attacks by identifying the central nodes of the network. The approach makes use of the PageRank and DeepWalk algorithms and Louvain community detection among other methods to derive riskiness scores for individual nodes and for communities. As an illustrative application, the authors consider the WannaCry ransomware-Bitcoin cash-out network and derive interesting insights.

In the third article entitled "High-Frequency News Sentiment and Its Application to Forex Market Prediction", the focus is on understanding how financial news information sentiment can be used in high-frequency Forex market prediction. The authors, Frank Z. Xing (Nanyang Technological University), Duc-Hong Hoang (Lund University), and Dinh-Vinh Vo (Lund University) propose to use a state-of-the-art NLP model, namely FinBERT, recently developed for financial sentiment analysis. To assess the performance of their proposed model, the authors compare the results to Naive Bayes classifier and a re-current neural network-based classifier among others, and the authors find a weak, but noticeable predictive power of the proposed model.

In the next article, "Predicting stock price and spread movements from news", Pontus Wistbacka (Han-ken School of Economics), Samuel Rönnqvist (Hanken School of Economics & University of Turku), Katia Vozian (Hanken School of Economics & Helsinki Graduate School of Economics) and Satchit Sagade (Goethe University Frankfurt & Leibnitz Institute for Financial Research SAFE) propose, similarly to the previous contribution, to use financial news on analyzing high frequency financial data. More specifically, their aim is to

HÍCSS

predict liquidity shocks. As an important outcome of the research, the results highlight the difficulties in specifying appropriate time spans in order to achieve high performance. After experimenting with various aggregation and filtering approaches and several advanced machine learning models, the authors conclude that future research should focus on the general problem setup rather than experimenting with various advanced machine learning models.

The fifth paper, "To Treat, or Not to Treat: Reducing Volatility in Uplift Modeling Through Weighted Ensembles", presents a study by Jannik Rößler, Roman Tilly and Detlef Schoder (University of Cologne) focusing on issues related to direct marketing, namely on appropriate customer segment identification. The authors consider the popular uplift modeling approach and aim to investigate the problem of high volatility typically associated to uplift modeling. They propose to use an ensemble approach combining three individual uplift-based models: two-model approach, a class transformation model, and a direct uplift model. In the tests, all models utilize Random Forests as the base prediction model. The empirical analysis on nine real-world data-sets show that the ensemble approach decreases the volatility significantly compared to individual models.

In the sixth and final accepted submission, "Hierarchical learning for option implied volatility pricing", Henry Han (Fordham University) proposes a new hierarchical learning procedure to be used in implied volatility pricing. As it is pointed out in the article, while this is a popular domain of allying various machine learning algorithms, in general it is challenging to achieve high performance. The learning approach introduced in the paper tackles the issues with traditional pricing methods as it boots the performance of individual models. The performance of the new approach is illustrated on a data-set of options traded in Nasdaq and NYSE.